

Expressing Visual Relationships via Language

Hao Tan, Franck Deroncourt, Zhe Lin,
Trung Bui, and Mohit Bansal

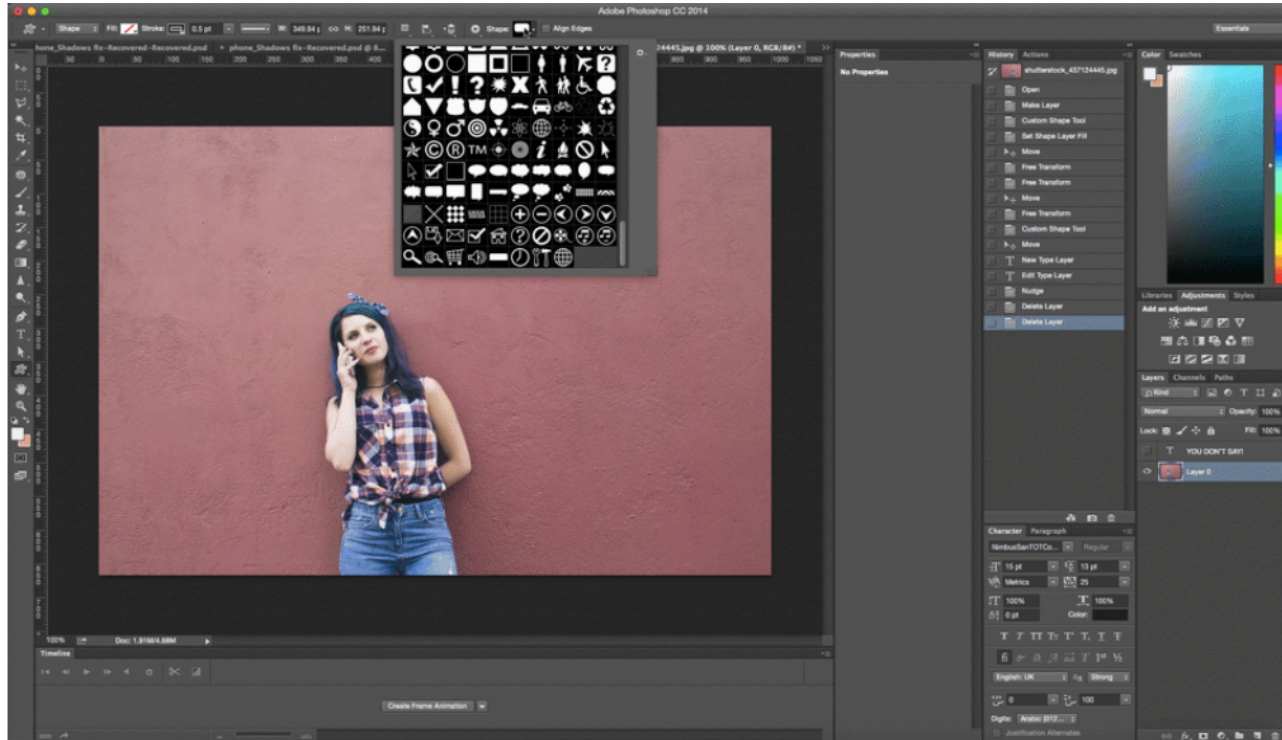
haotan, mbansal@cs.unc.edu
{deronco, zlin, bui}@adobe.com

UNC Chapel Hill & Adobe Research

Image Editing

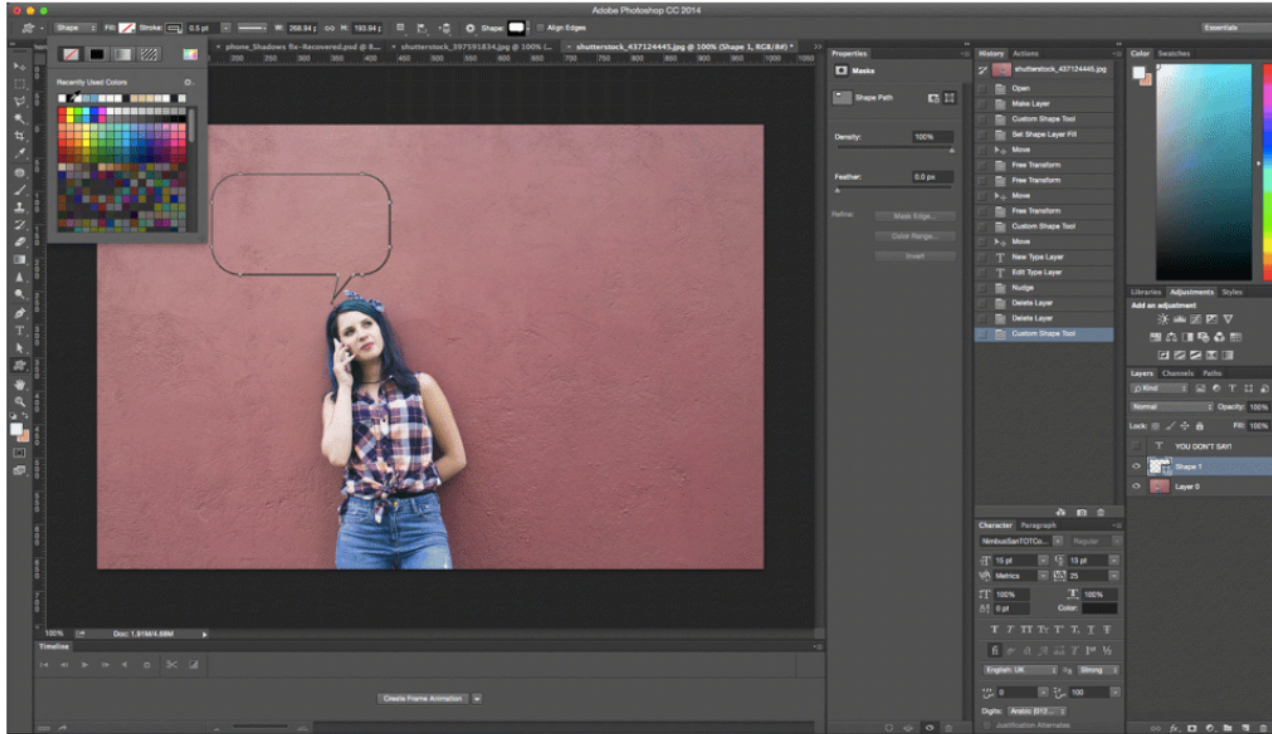


Add a sword and a cloak to the squirrel.



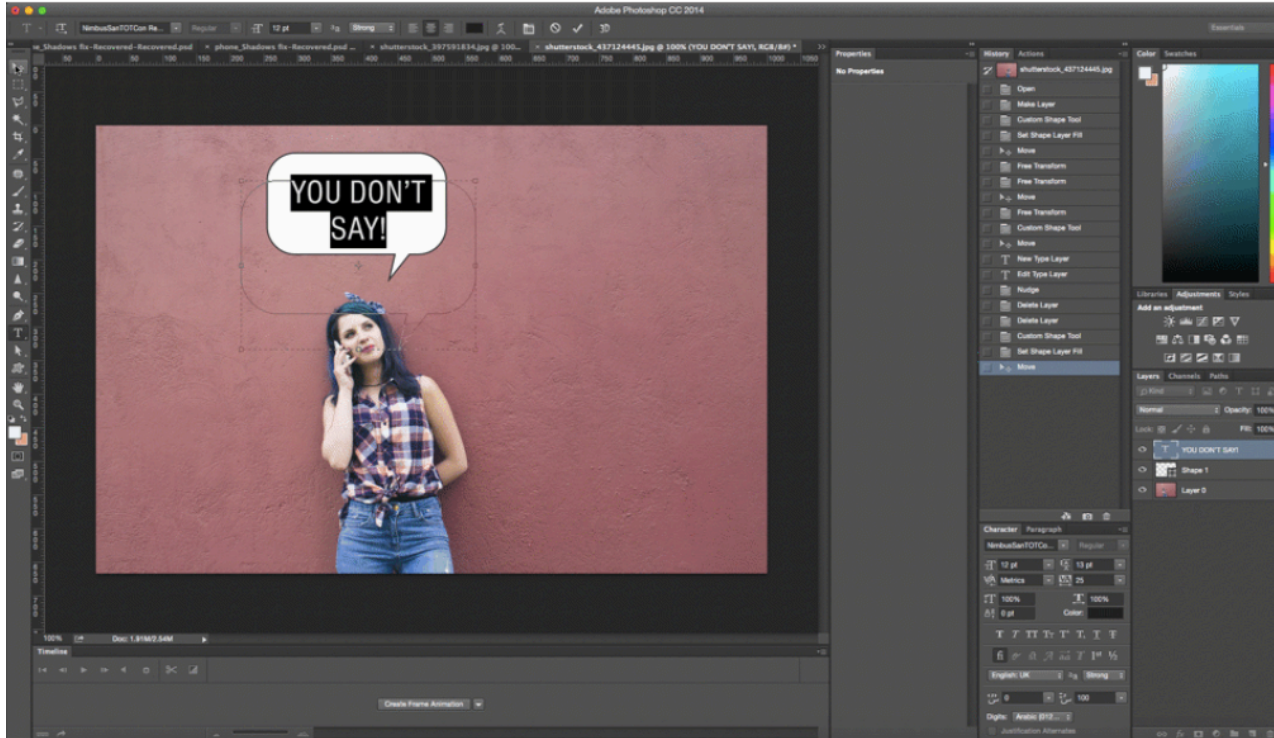


Photoshop



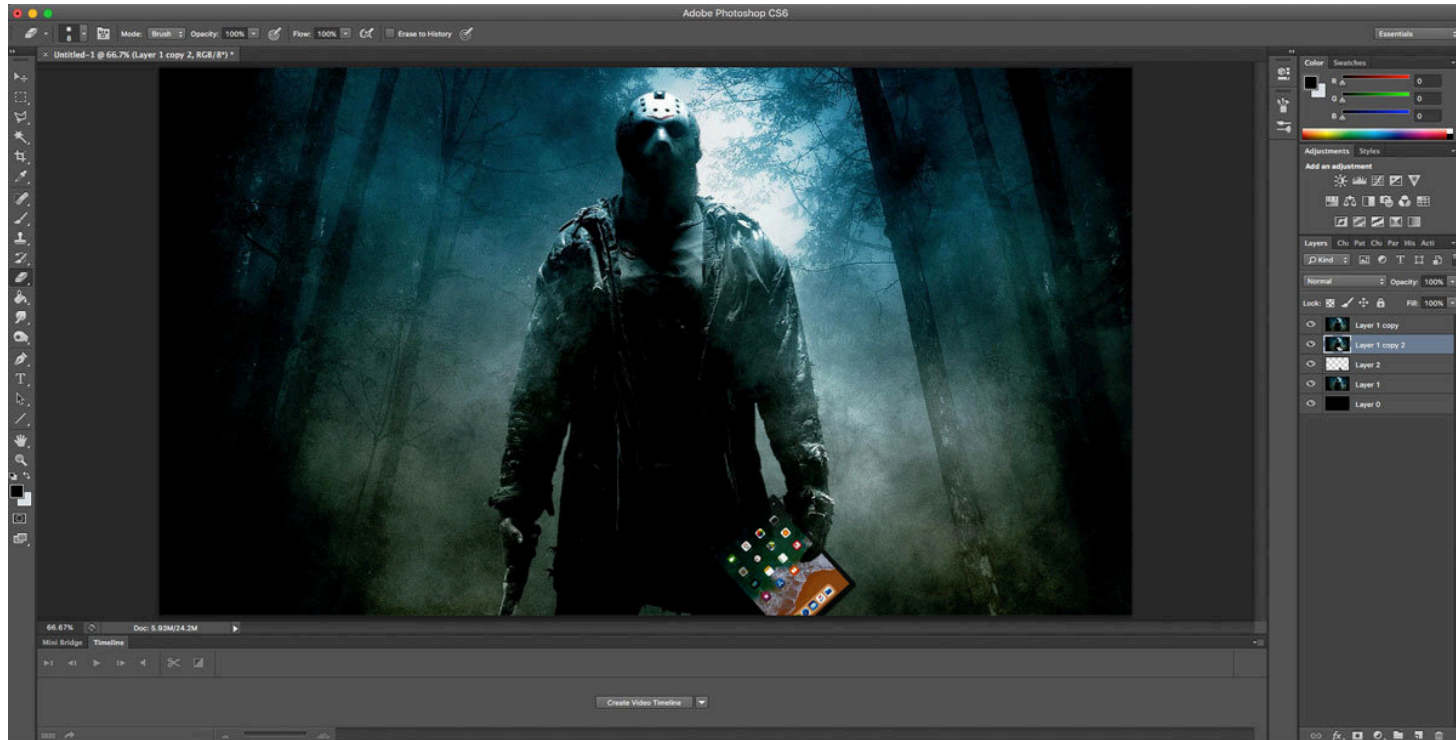


Photoshop





Photoshop



Could we build models that use **human language** to automatically edit images?

Image Editing Pipeline



Add a sword and a
cloak to the squirrel.



Me

Image Editing Pipeline (Final Goal)

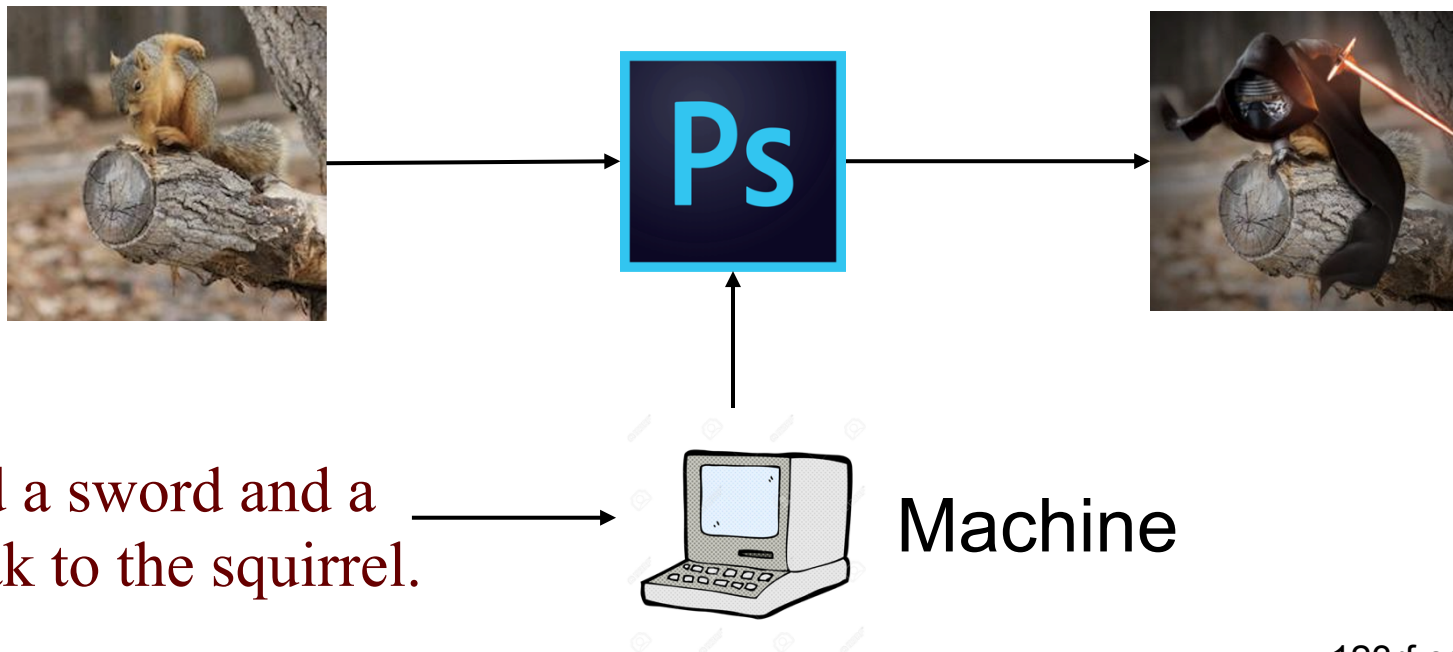
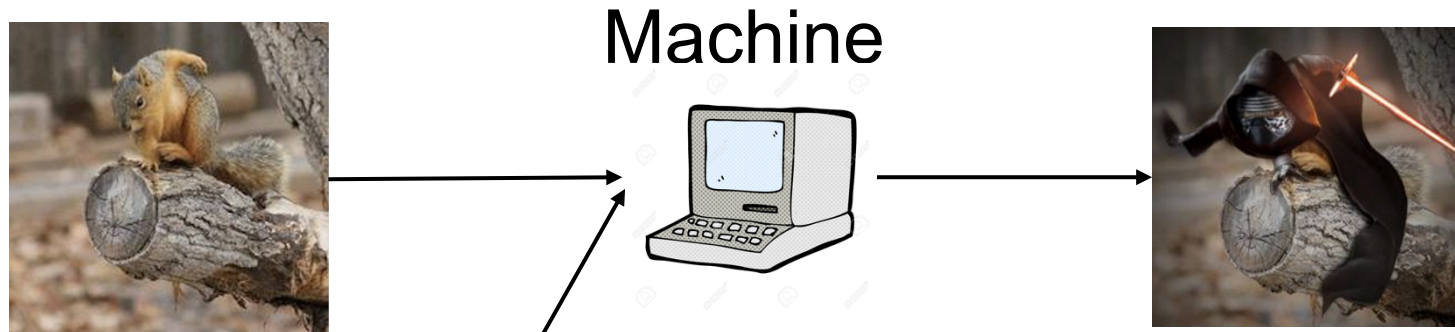


Image Editing Pipeline (Final Goal)



Add a sword and a
cloak to the squirrel.

ImgEdit (Image Editing Request) Corpus



Source Image



Editing Request

Add a sword and a
cloak to the squirrel.

Target Image



ImgEdit (Image Editing Request) Corpus



Source Image



Target Image



Editing Request

Add a sword and a
cloak to the squirrel.

Collected from Internet

Annotated by Human

Collect Images



Reddit
(r/Photoshop
Request/)

Collect Images

Reddit
(r/Photoshop
Request/)



Collect Images

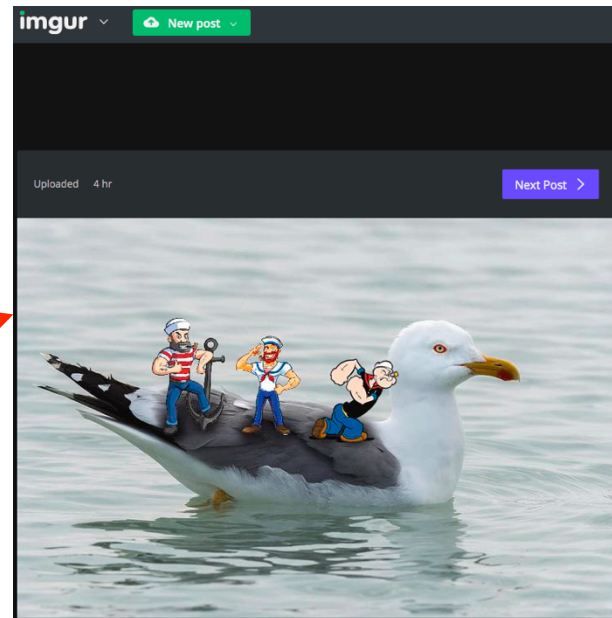
Reddit
(r/Photoshop
Request/)



5 Comments Share Save Hide Report 71% Upvoted

↑ Francophile_45 Score hidden · 4 hours ago
↓ My attempt
↑ Blood_Eagle_IX Score hidden · 4 hours ago
↓ Lol.. thanks! That's pretty good.

↑ imthatguyhere Wizard Score hidden · 4 hours ago
↓ Here's a thing:
<https://i.imgur.com/iGvm0qd.png>
↑ Blood_Eagle_IX Score hidden · 4 hours ago
↓ Thank you! Very.. singing in the rain quality for some reason.... it reminds me of a movie from the 40s



<https://www.reddit.com/r/PhotoshopRequest/> 15

Collect Images

Reddit
(r/Photoshop
Request/)

↑ Posted by u/Blood_Eagle_IX 5 hours ago
4
↓ [RANDOM] Please let me try this again mods. Can someone please add a bunch of SAILORS riding on the back of this seagull. It's a running joke for a group. Thank you.



5 Comments Share Save Hide Report

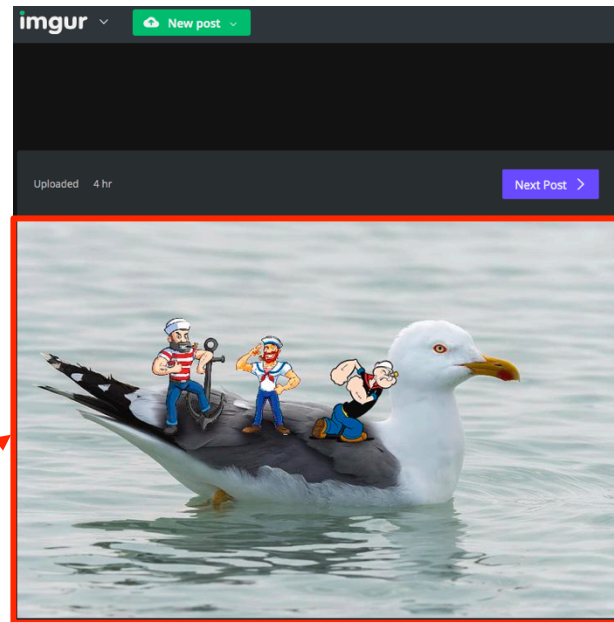
71% Upvoted

↑ Francophile_45 Score hidden · 4 hours ago
↓ [My attempt](#)

↑ Blood_Eagle_IX Score hidden · 4 hours ago
↓ Lol.. thanks! That's pretty good.

↑ imthatguyhere Wizard Score hidden · 4 hours ago
↓ Here's a thing:
<https://i.imgur.com/iGvm0qd.png>

↑ Blood_Eagle_IX Score hidden · 4 hours ago
↓ Thank you! Very.. singing in the rain quality for some reason.... it reminds me of a movie from the 40s



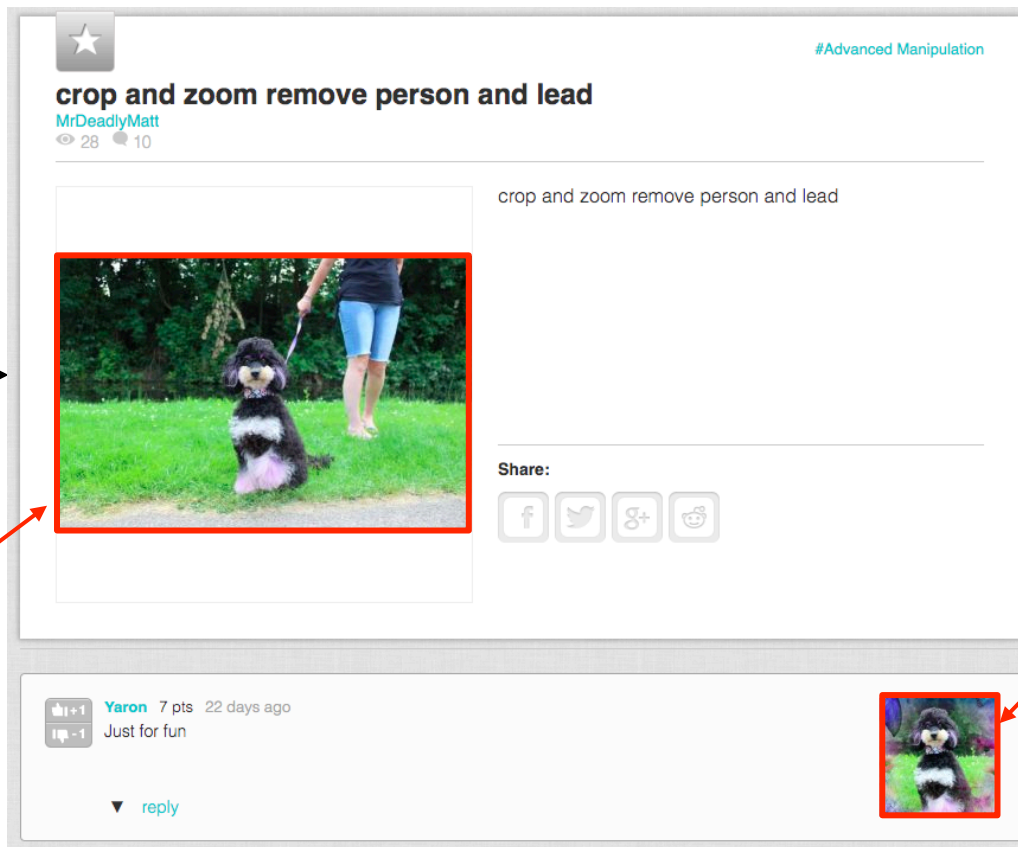
**Source
Image**

**Target
Image**

Collect Images

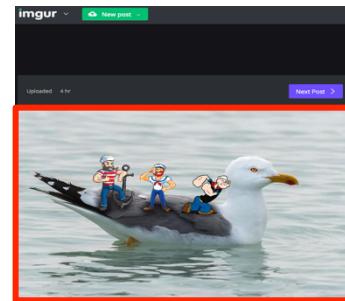
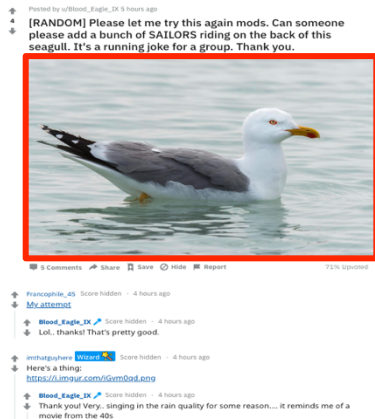
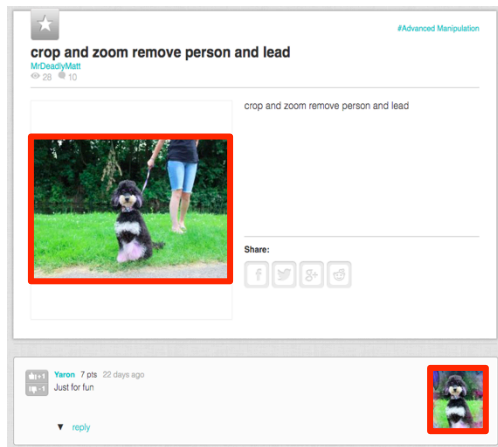
Zhopped →

Source
Image



Target
Image

Collect Images



Zhopped

+

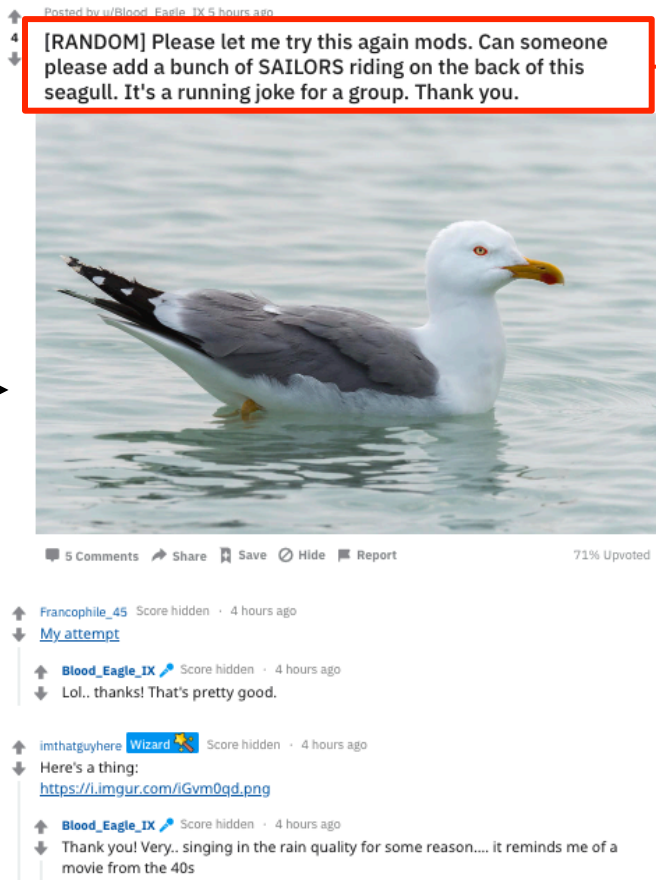
Reddit
(r/Photoshop
Request/)

=

12K image pairs

Collect Editing Requests

Reddit →
(r/PhotoshopRequest/)



[RANDOM] Please let me try this again mods. Can someone please add a bunch of SAILORS riding on the back of this seagull. It's a running joke for a group. Thank you.

Collect Editing Requests

Reddit

(r/PhotoshopRequest/)



[RANDOM] Please let me try this again mods. Can someone please add a bunch of SAILORS riding on the back of this seagull. It's a running joke for a group. Thank you.

Specifications are too noisy!!

Collect Editing Requests

Source
Image



Target
Image



Collect Editing Requests

Source
Image



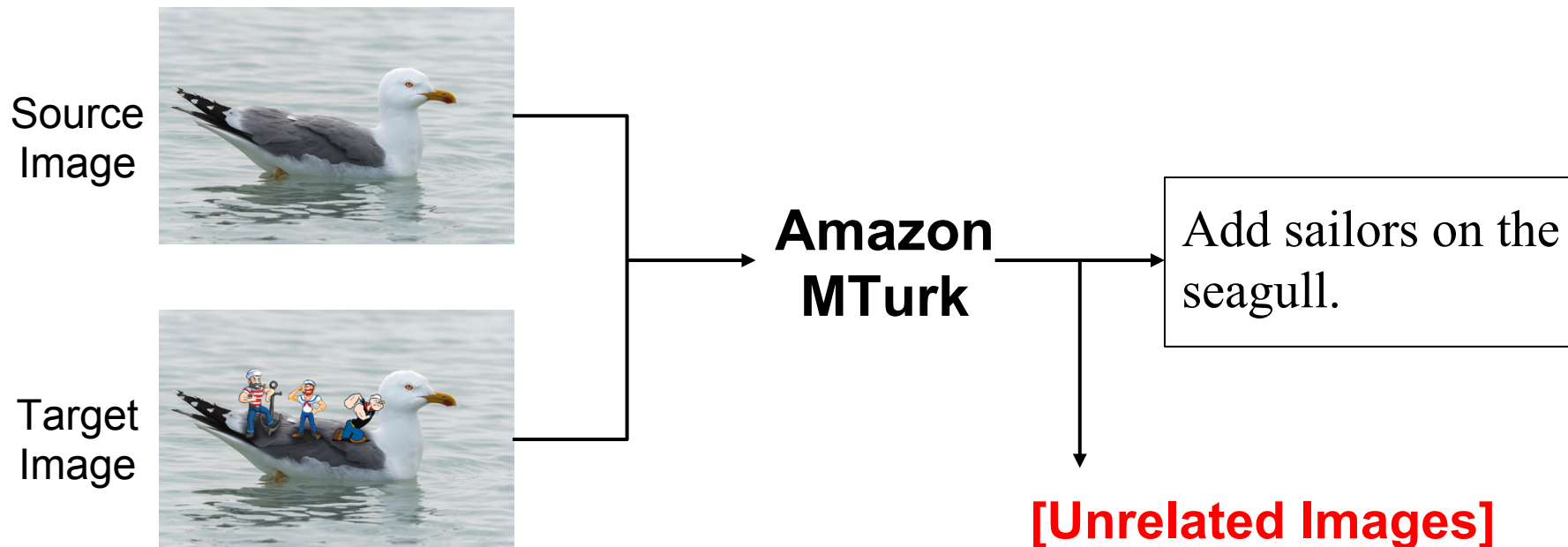
Target
Image



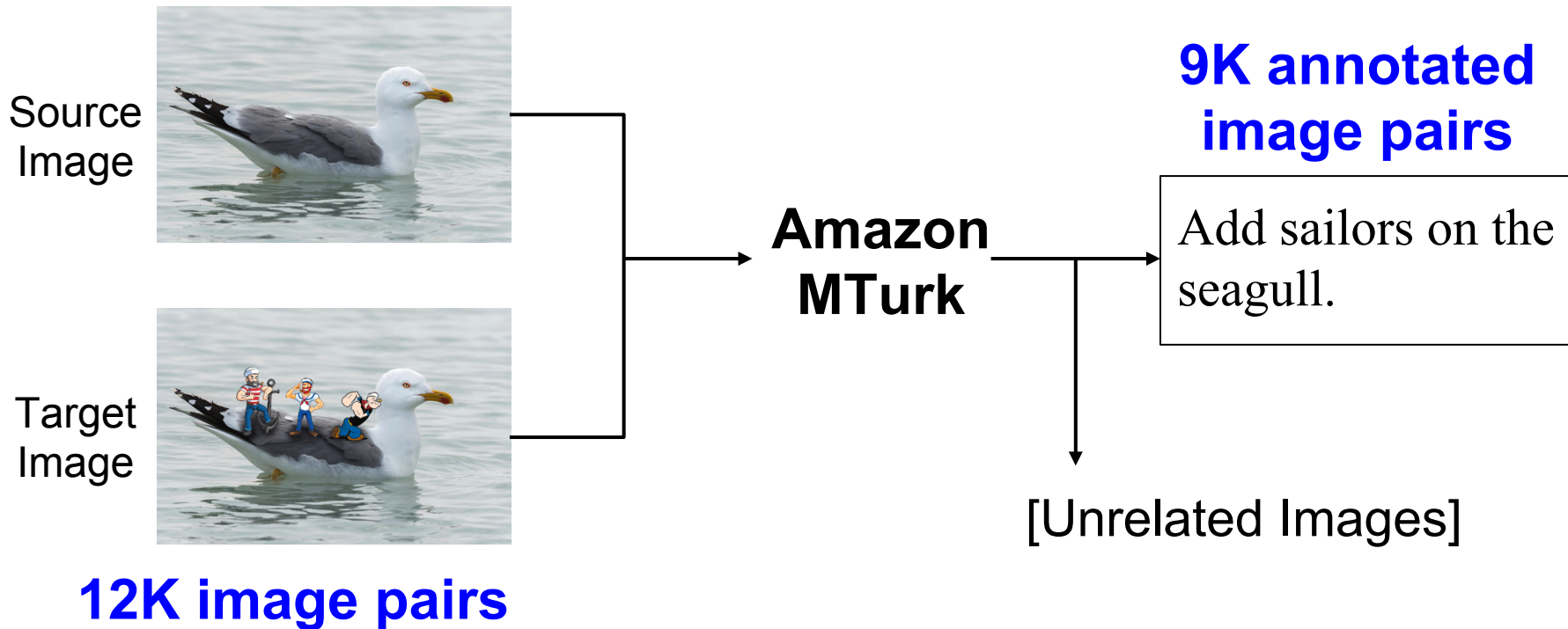
**Amazon
MTurk**

Add sailors on the
seagull.

Collect Editing Requests



Collect Editing Requests



Quality Control (Second Stage)

Source
Image



Target
Image



Request

Add sailors on the
seagull.



Correct

Incorrect

Quality Control (Second Stage)

Source
Image



Target
Image



Request

Add sailors on the seagull.

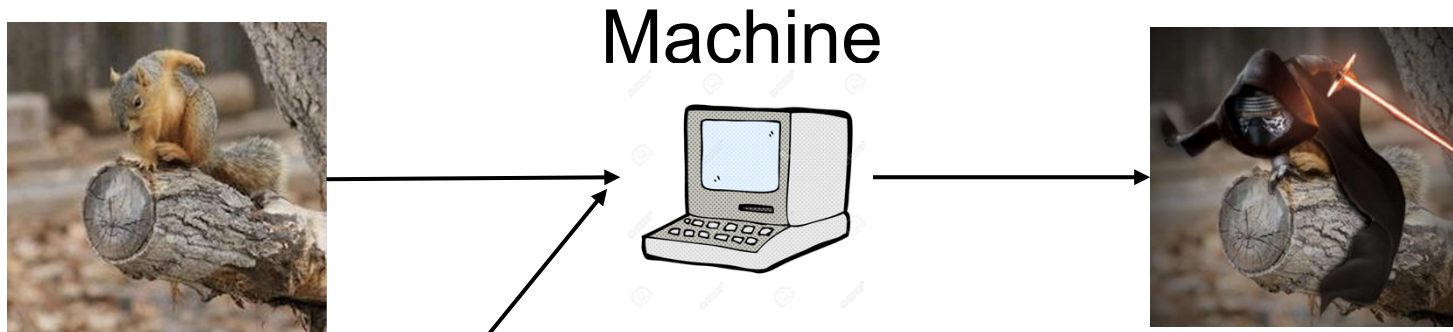


4K image pairs

Correct

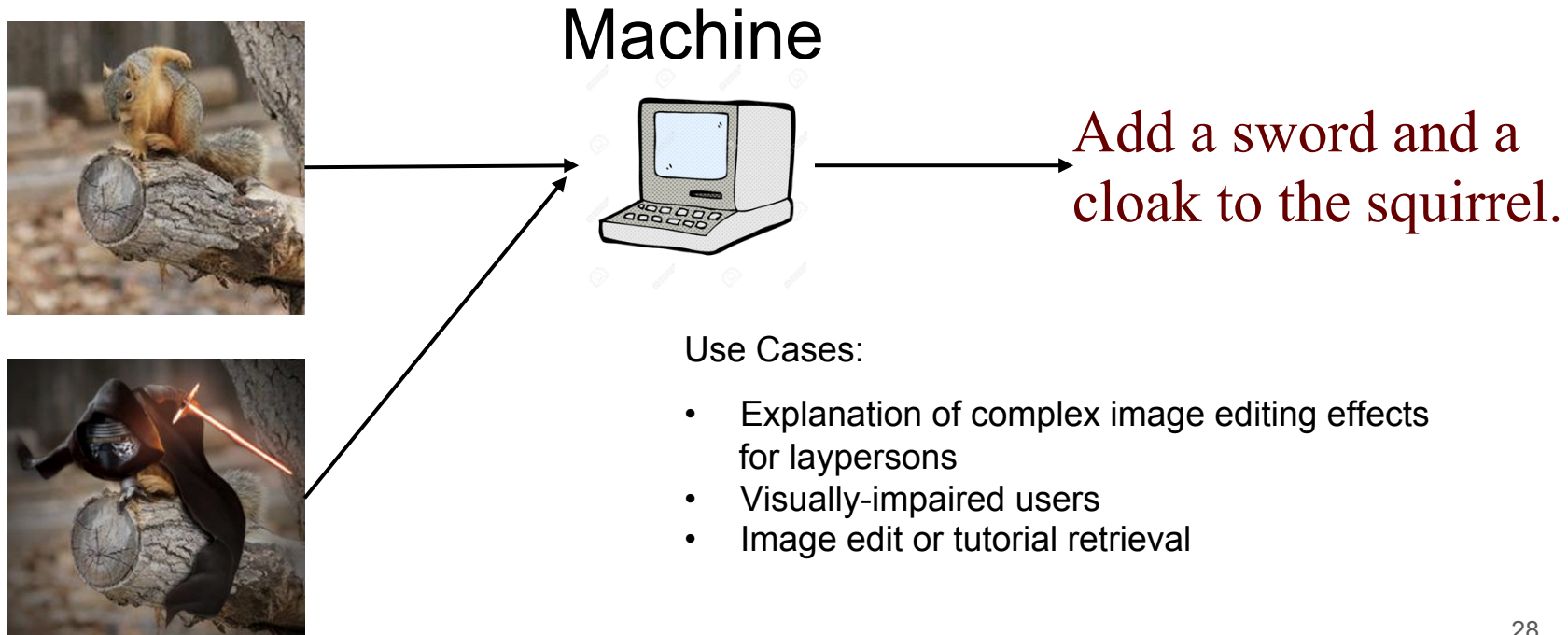
Incorrect

Task: Editing Request Execution



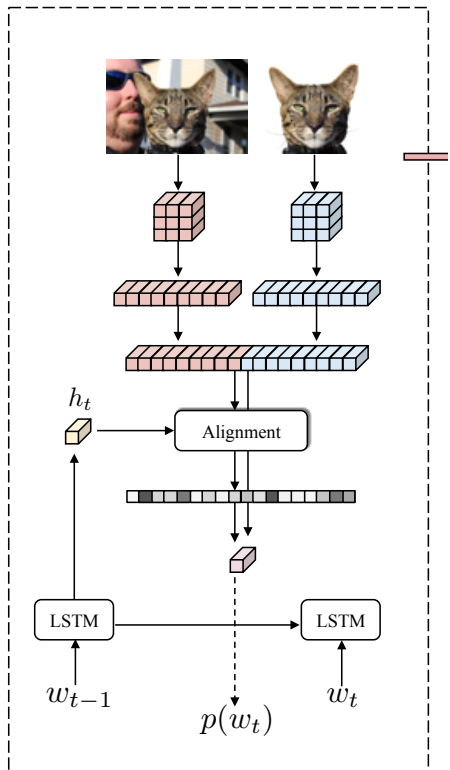
Add a sword and a
cloak to the squirrel.

Task: Editing Request Generation

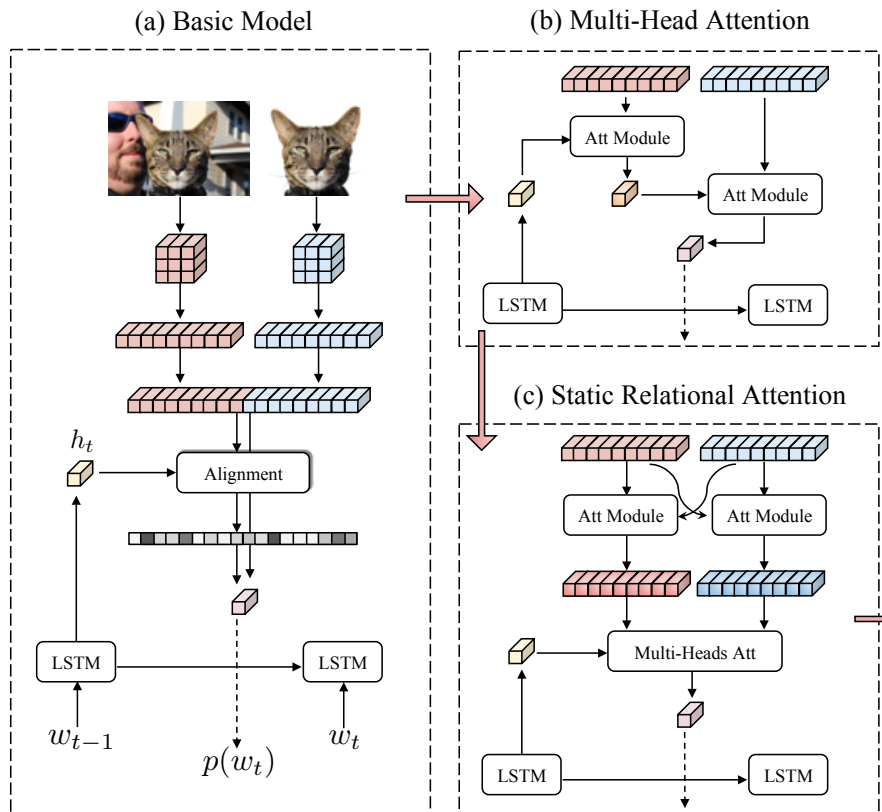


Basic Captioning Model

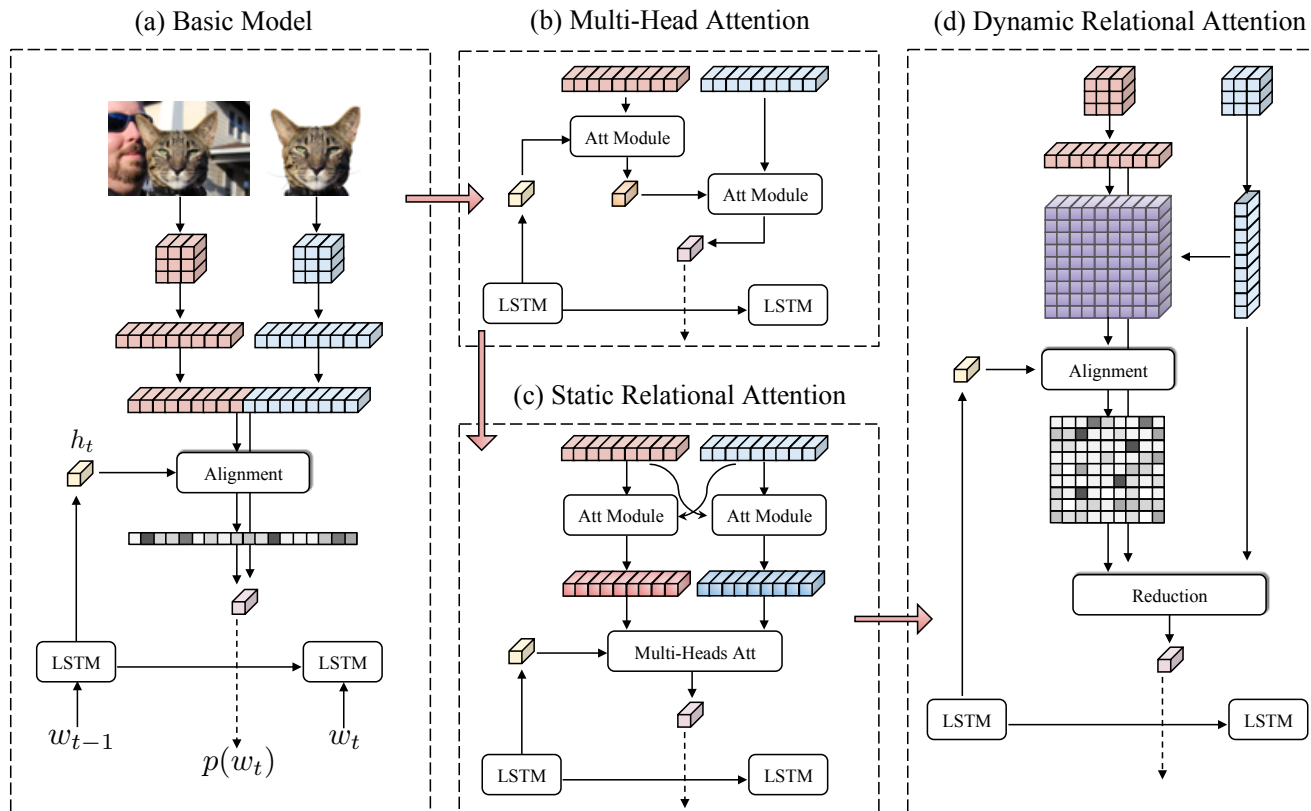
(a) Basic Model



Basic Captioning Model



Full Captioning Model



3 Evaluation Datasets: ImgEdit



Remove the man from the image.

3 Evaluation Datasets: Spot-the-Diff



The blue truck is no longer there.

A car is approaching the parking lot from the right.

3 Evaluation Datasets: NLVR2



Each image shows a row of dressed dogs posing with a cat that is also wearing some garment.

3 Evaluation Datasets

Ours (Image Editing Request)



Add a sword and a cloak to the squirrel.

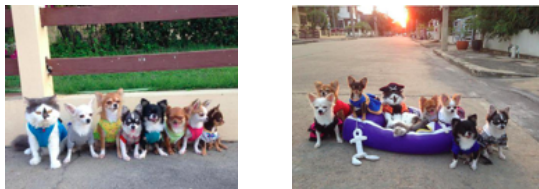
Spot-the-Diff



The blue truck is no longer there.

A car is approaching the parking lot from the right.

NLVR2 Captioning



Each image shows a row of dressed dogs posing with a cat that is also wearing some garment.

Convert

NLVR2 Classification



Each image shows a row of dressed dogs posing with a cat that is also wearing some garment. → True



In at least one of the images, six dogs are posing for a picture, while on a bench. → False

Evaluation Methods

Phrase-based Metrics:

BLEU, **CIDEr**, METEOR, ROUGE-L

Human Evaluation:

Pairwise Comparison

Phrase-based Metric: CIDEr

ImgEdit



21.6 \rightarrow **26.4**
(+4.8)

NLVR2



43.4 \rightarrow **46.4**
(+3.0)

Spot-the-Diff



26.3 \rightarrow **35.3**
(+9.0)

Human Evaluation: Winning Rate

ImgEdit



11% vs. **24%**

NLVR2



24% vs. **37%**

Spot-the-Diff

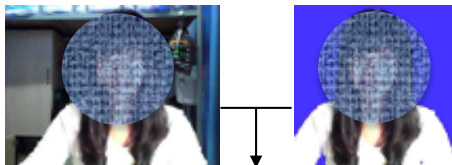


22% vs. **37%**

Generated Results

Image Editing Request

Positive
Examples



change the background to blue

Negative
Examples

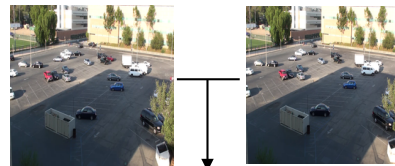


add a filter to the image

Spot-the-Diff



the person in the white shirt is gone

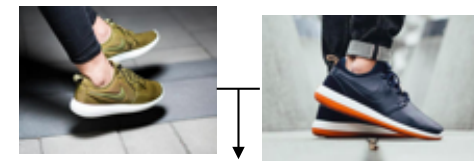


the black car in the middle row is gone

NLVR2



there is a bookshelf with a white shelf in one of the images .



the left image shows a pair of shoes wearing a pair of shoes .

Thank you!

Code at:

<https://github.com/airsplay/VisualRelationships>

UNC Chapel Hill & Adobe Research